DOCUMENT RESUME

ED 141 409 TM 006 374

AUTHOR Masters, James R.; Weaver, Waldo G.
TITLE The Development of a Student Observation of Teachers Instrument for Use in High Schools.
PUB DATE Apr 77
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, New York, April 5-7, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Behavior Rating Scales; Classroom Observation Techniques; Effective Teaching; Factor Analysis; *High School Students; *Secondary School Teachers; Student Attitudes; *Student Evaluation of Teacher Performance; Teacher Evaluation; Teaching Methods; Teaching Techniques; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS *Student Observation Teachers Teaching Techniques

ABSTRACT

The Student Observation of Teachers and Teaching Techniques instrument (StOTT) was developed by high school students who served on the Student Advisory Board to the Pennsylvania Department of Education. The StOTT instrument consists of 32 items. Each of the first 29 is included in one of the subscales: Teacher-Student Relations, Grades and Testing, Materials, Teacher Personality, and Teaching Methods and Techniques. The final three items ask students to give their opinions about whether or not the teacher they are rating has any distracting personal mannerisms, to give the reason they took the teacher's course, and to tell whether they would recommend the course to another student. As reported in the Educational Quality Assessment manual (Hertzog, 1976) this instrument has evidenced an internal consistency reliability of .86. As evidences of validity significant correlations have been found with such variables as the number of books read by students, the number of days they were truant, and their desires to quit school (Masters, Shannon and Reardon, 1975). A copy of the Student Observation of Teachers test (version 2) is appended. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141409

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DEVELOPMENT OF A STUDENT OBSERVATION OF TEACHERS INSTRUMENT FOR USE IN HIGH SCHOOLS

James R. Masters
Waldo G. Weaver

Pennsylvania Department of Education

Presented at the Annual Meeting of the National Council on Measurement in Education

New York

April 1977

TM006 374

Summary

This study was performed to investigate the reliability and validity of an instrument designed to provide student feedback to high school teachers about their methods. The instrument examined, the Student Observation of Teachers and Teaching Techniques Instrument (StOTT), was developed by high school students who served on the Student Advisory Board to the Secretary of Education of Pennsylvania.

Included in the study were 925 high school students and 36 teachers. Other variables were collected along with StOTT ratings as a means of investigating their influence.

Support was gathered in the study for scoring the StOTT in terms of five subscales. Student ratings of teachers were found to be highly stable over an interval of one month. The instrument seemed to be capable of measuring ways in which students felt that their teachers differed from each other. The finding that student ratings appeared to be influenced by such factors as student attitudes toward school and teachers and student success in school indicated that the instrument should not be used to evaluate and compare teachers. However, the instrument seemed to be a valuable one for use within individual teachers' classrooms.

Introduction

Probably most high school teachers have at one time or another made some attempt to gather student opinions about their teaching methods. Teachers who have used student feedback instruments would generally agree that information collected in this way is valuable to them. Hayes, Keim and Neiman (1967), for example, reported that over 90 per cent of the high school teachers in their study believed that useful information was obtained from the student feedback instrument they were asked to use.

But, how accurately can students assess teaching methods? Are student ratings of a teacher's techniques stable over time? Do extraneous factors, such as a student's grade in a teacher's course, influence his or her ratings of the teacher?

Questions such as these have been asked repeatedly in recent years by investigators concerned with the reliability and validity of instruments designed for use at the college level. They have been asked much less often about instruments designed to be used by high school teachers, possibly because few attempts have been made at this level to employ such instruments as vehicles for evaluating and comparing teachers.

Research which has been carried out at the high school level has produced conflicting results. On the positive side a number of investigators [e.g., Hayes, Keim and Neiman (1966, 1967), Davidoff (1970), DeAmico (1973)] have reported results which indicated that student ratings of teachers were stable over time. Both in the study carried out by Hayes, Keim and Neiman (1966) and in that performed by Veldman and Peck (1967) it was found that student ratings of teachers had some agreement with ratings given these same teachers by their supervisors. Thompson (1974) uncovered no significant relationship between ratings of high school teachers and their students' sex, grade level, grade-point average, expected course grade or degree of absenteeism. Shaw (1973) concluded that ratings of high school teachers were not influenced by the teacher's sex, academic degrees or years of teaching experience or by students' grade-point average.

On the negative side, however, a number of studies have indicated that student ratings of high school teachers are influenced by factors other than teachers' behaviors. For example, Smith and Brown (1976) concluded that students' general attitudes toward teachers, their enjoyment of the subject matter and the grade they expected to receive all influenced ratings given a teacher. Jackson and Fuller (1966) reported that female students perceived teachers as more confident or poised than did male students. They also found that the socioeconomic status of students and teachers seemed to influence ratings received by teachers. Veldman and Peck (1967) found that student ratings of teachers varied as a function of the subject area taught. Teachers teaching physical education, home economics or business were found to receive the highest ratings. Those teaching foreign languages, mathematics or science received low ratings.

The present study was carried out to examine the reliability and validity of an instrument developed to provide student feedback to high school teachers about their methods. Many of the same questions addressed in past investigations of student feedback instruments were asked as a means both of examining closely a specific instrument and of adding to existing knowledge about the use of such instruments in high schools.

The Student Observation of Teachers and Teaching Techniques Instrument

Description of the Instrument

The instrument examined in the study, the Student Observation of Teachers and Teaching Techniques instrument (StOTT), was developed by high school students who served on the Student Advisory Board to the Pennsylvania Department of Education. This Board, composed of eleventh and twelfth grade student representatives from all parts of Pennsylvania, was formed to advise the Secretary of Education of student opinions about Pennsylvania's educational system.

The StOTT instrument consists of 32 items. Each of the first 29 is included in one of five subscales. The final three items ask students to give their opinions about whether or not the teacher they are rating has any distracting personal mannerisms, to give the reason they took the teacher's course and to tell whether they would recommend the course to another student.

The first 29 items are positive statements about a teacher's classroom behaviors (e.g., The teacher is willing to help you when you need it). For each of these a five-point scale is provided for students to mark their responses. The first, third and fifth points of the scale are defined for the student. These are, in the majority of cases, temporal (i.e., Always, Sometimes, Never) but for certain items a somewhat different scale is employed. An example of such an item is the following:

The material on tests was covered by the teacher prior to the test.

5	4	3	2	1
Thoroughly		Moderately		Not at All

When examining the results of a class' rating of him or her a teacher is encouraged to look at both the results for individual items and the means obtained on each of the subscales. These five subscales are Teacher-Student Relations (6 items), Grades and Testing (6 items), Materials (3 items), Teacher Personality (3 items) and Teaching Methods and Techniques (11 items).

Development of the Instrument

The development of the StOTT was begun in October of 1973 at the first meeting of the Student Advisory Board. A Tenure/Teacher Evaluation Committee was formed to study how students might help improve teaching. A decision was made by this Committee to find or develop a student observation of teachers instrument of proven worth. After examining a number of questionnaires in use in the United States, the Committee, assisted by their Department of Education advisers, began work on the development of their own instrument. This instrument, if proven to be of quality, would be made available to high school teachers who desired to obtain feedback from their students about the strengths and weaknesses of their methods.

By the end of the 1973-74 school year questions for an instrument had been written. These were taken into high school classrooms in nine school districts to obtain student opinions about them. Students were asked to examine the clarity and appropriateness of each question. Teachers and administrators were also asked to examine the instrument at this time. Revisions in wording and the deletion of one question resulted from this tryout.

During the 1974-75 school year the revised instrument was taken into classrooms again. This time students both reacted to the appropriateness and clarity of the questions and indicated their observations about their teachers. Again, the wording of a number of questions was changed. Some evidences of reliability and validity were found in this tryout.

Since the instrument now showed promise of being a useful one it was decided that a large-scale study was needed to examine further its reliability and validity. The study described in this paper was designed for this purpose.

Methodology of the Study

Sample

The study was carried out during the 1975-76 school year. In all, 36 teachers participated, some of whom were rated by more than one class. A total of 925 tenth, eleventh and twelfth grade students from 45 classes and five schools were included in the study. Two of these schools were suburban, two were rural and one was a vocational-technical school.

Design

As a means of obtaining as much information as possible from the tryout without taking up a great deal of any teacher's instructional time, three distinct groups were used. By employing different procedures and measurement instruments with each group it was possible to ask a variety of research questions about the StOTT.

Group 1 consisted of 18 teachers whose classes filled out the StOTT, waited a month, then filled out the instrument again. In addition, these students answered a questionnaire giving information about themselves, including their sex, grade level, curriculum, reason for taking the course and average grade both for the teacher's course and for their major courses. This group was used to examine both the stability over time of student ratings of the same teacher and the influence of extraneous factors upon student ratings of a teacher.

Group 2 consisted of eight teachers, each of whom asked two of their classes to rate them. Both classes were of the same grade level and were taking the same subject from the teacher. Information was gathered in this way on the extent to which ratings of a teacher were a function of the particular class chosen for the rating task. Such information would be useful to teachers in interpreting the results of StOTT administrations.

One of the classes of each Group 2 teacher was asked also to fill out a forced-choice version of certain of the items of the StOTT without looking at the StOTT ratings they had given a teacher. This instrument (shown in the Appendix) was made up of four items from the Teacher-Student Relations subscale and four items from the Methods subscale. Each of the Teacher-Student Relations items was paired with each of the Methods items, producing a total of 16 pairs. Students responding to the instrument were asked to determine which statement of each pair was more true of the teacher they were rating. A student's score on the instrument was the number of times he or she had chosen a Teacher-Student Relations item over a Methods item. For the same 16 pairs the same student's item scores on the StOTT were examined. The number of times he or she had given a higher rating to a Teacher-Student Relations item than to a Methods item was computed (e.g., a rating of 5 on item 1 and a rating of 4 on item 23 would be counted in a similar way to choosing item 1 over item 23 in a forced-choice pairing). Correlations of students' scores obtained on the two instruments provided an index of students' abilities to make thoughtful judgments when rating teachers on specific items of the StOTT. High correlations would indicate that, at least at the time of the rating task, students were capable of making rather fine judgments which had meaning to them.

The other class of each Group 2 teacher responded to the instrument used in Pennsylvania's statewide assessment program (Educational Quality Assessment) to measure eleventh grade students' Interest in School and Learning. The 37-item instrument is made up of three subscales, measuring attitudes toward Learning, toward School and toward Teachers. As reported in the Educational Quality Assessment manual (Hertzog, 1976) the instrument has evidenced an internal consistency reliability of .86; the reliabilities of the Learning, School and Teachers subscales have been found to be, respectively, .60, .82 and .70. As evidences of validity significant correlations have been found with such variables as the number of books read by students, the number of days they were truant and their desires to quit school (Masters, Shannon and Reardon, 1975). Correlations of student scores on the Interest in School and Learning instrument with their StOTT ratings were used to determine how much influence students' general attitudes toward school and teachers had upon their ratings of an individual teacher.

Group 3 consisted of nine classes, each of which rated two of their teachers. Comparisons of the mean ratings given two teachers by the same class were carried out to learn if, in fact, students were capable of using the StOTT to pick out what they felt were specific strengths and weaknesses of their teachers' methods. If classes were found to give similar ratings to each of their teachers, the worth of the StOTT instrument would have to be seriously questioned.

Results

Reliability and Independence of the StOTT Subscales

Before an examination of other questions about the StOTT could be attempted some assurance was needed that its subscales were reliable and relatively independent of each other. The subscales were formed on a conceptual basis, by placing items of like content on the same subscale. Little empirical support had been gathered for this structuring.

Therefore, an item analysis was carried out using data from all 925 students who had filled out the StOTT. Table 1 contains the internal consistency reliability coefficients (coefficient alphas) computed for each of the subscales. As shown, reliabilities ranged from .51 (Materials) to .80 (Teacher-Student Relations). Each item of the instrument correlated more highly with its own subscale than with any other subscale. The item analysis results, then, generally supported the five-subscale structure. The reliability of the Materials subscale, however, was somewhat low. Since the items included on it did correlate with each other the relatively low reliability obtained would seem to indicate that more items than the three employed are needed to provide a general measure of this area.

Also shown in Table 1 are the mean item ratings given teachers in each of the subscale areas. (A scale ranging from 1 to 5 was employed in scoring each item.) The highest ratings were given for the items of the Grades and Testing subscale and the items of the Materials subscale. The lowest ratings were given in assessing Teaching Methods and Techniques.

Table 1

Reliabilities and Mean Item Ratings Obtained for Each StOTT Subscale

Subscale	No. of Items	Mean Item Rating	Coefficient Alpha
Teacher-Student Relations	6	4.12	.80
Grades and Testing	6	4.27	.65
Materials	3	4.28	.51
Teacher Personality	3	4.13	.71
Teaching Methods and Techniques	11	3.68	.79

Table 2 shows the correlations obtained among the five subscale areas. The highest relationship between two subscales, .70, was found for the correlation between Teacher Personality and Teacher-Student Relations. The results summarized by Table 2, then, also supported the use of the five subscales, since each appeared to provide information which was somewhat unique.

Table 2

Correlations Among the StOTT Subscales

	Teacher-Student Relations	Grades/ Testing	Materials	Personality
Grades/Testing	.54			
Materials	.38	.42		
Personality	.70	.50	.30	
Methods	.62	.63	.47	.64

Thus, support was gathered for scoring the StOTT in terms of the conceptually based subscales. Scores on these could therefore be used in further investigations of the properties of the instrument.

Stability of Ratings Over Time

Included in Table 3 are correlations summarizing the degree of stability found for student ratings over an interval of one month. Results for the total group of 334 students who participated in this phase of the study and for two subgroups are shown. It was believed possible that students experiencing learning difficulties in school might, because of lack of motivation and inability to read well, produce ratings which had a degree of instability. Therefore, a comparison was made of the stability of ratings given by students averaging grades of A or B in their major courses with that of students averaging grades of C and below. Finally, mean ratings given each of the 18 teachers on the two occasions by their classes were correlated as an index of the stability of these.

As Table 3 indicates student ratings of the same teacher did evidence a great deal of stability over the one month interval. For the total group of students stability coefficients for four of the subscales were greater than .70; again, the results for the Materials subscale were lowest. Students earning grades of C and below in school did not appear to give ratings which were any less stable than were those of students earning grades of A or B. Especially impressive were the stability coefficients for the mean ratings given each teacher by their classes. For all five subscales, coefficients of .75 or greater were computed and for three subscales these coefficients were greater than .90.

Table 3

Stability Coefficients for the StOTT

Group	No. of Students	Teach.-Stud. Relations	Grades/ Testing	Materials	Person-ality	Methods
Students Averaging Grades of C or Below	109	.80	.67	.60	.78	.70
Students Averaging Grades of A or B	208	.79	.79	.61	.72	.73
ALL STUDENTS	334	.79	.76	.60	.75	.71
Mean Ratings Given to 18 Teachers		.93	.92	.79	.91	.76

Ability of Students to Make Meaningful Decisions

As described in the Design section certain of the StOTT items were included in a forced-choice instrument. Scores on this instrument were correlated with scores obtained for the same items of the StOTT. Scoring was accomplished such that scores on both instruments reflected the degree to which a teacher was rated higher in Teacher-Student Relations than in Methods. Table 4 contains correlations computed to examine the relationship between scores obtained in the two ways. Results for each of the eight classes and for the total group are shown.

As is evident in Table 4, scores obtained on the two instruments agreed well with each other for six of the eight classes and for the total group (one-tailed tests at the .05 level were employed). Students, therefore, did appear in general to be making decisions about StOTT item ratings which had meaning to them. When confronted with the same items in a different response format students gave responses which were highly similar to those they had given using the five-category StOTT scales.

Table 4

Correlations Between StOTT Scores and Forced-Choice Instrument Scores

Class	Number of Students	Correlations
A	26	.23
B	19	.62**
C	28	.81**
D	21	.72**
E	9	.59*
F	23	.27
G	26	.87**
H	19	.78**
Total Group	171	.70**

*Significant at beyond the .05 level
**Significant at beyond the .01 level

Ratings of Two Teachers by the Same Class

Table 5 shows rather conclusively that the same class was capable of using the StOTT to pick out ways in which they believed two of their teachers differed. Each of the nine classes rated one of their teachers significantly higher than the other one on at least one of the five subscales (correlated t values were computed for these comparisons; two-tailed tests at the .05 level were performed).

Of especial interest are the results shown for the first two classes. These classes rated one of their teachers higher on the Grades and Testing subscale and the other teacher higher on the Personality subscale. This would seem to mean that these classes were not responding to some global feeling about their teachers but were, in fact, attempting to look at each StOTT item as a rating task in and of itself.

Ratings of the Same Teacher by Two Classes

Table 6 summarizes the results found in comparing the ratings given the same teacher by two classes. Two-tailed analyses of variance were carried out; F values for the comparisons are shown.

The results contained in Table 6 indicate that it is possible for a teacher to be given different ratings by two similar classes. This occurred for three of the eight teachers. The fact that differences were found for only three teachers would seem to suggest that the question asked is a complex one. The finding of differences may be a function of the particular teachers and classes involved. A much larger number of teachers than eight are needed to investigate this question more fully.

Influence Upon Ratings of Extraneous Factors

Tables 7-10 summarize the results uncovered in investigating the influence upon ratings of variables other than teachers' classroom behaviors. Tables 7 and 8 contain correlations between ratings given by students and both student and teacher characteristics. In Table 7 are shown correlations computed using the responses of the total group of 334 students utilized in this phase of the study. To construct Table 8, means were computed on each student characteristic variable for each of the 18 classes. These means, along with the Teacher Sex variable, were correlated with the mean ratings given each teacher on the five StOTT subscales.

Many of the correlations shown in Table 7 are significant. However, their magnitude is not great. Cohen (1969) would describe them as indicative of effects of small to medium size. Those variables which correlated significantly with ratings would be capable of predicting only a small proportion of their variance.

Table 5

Comparisons of Ratings Given Two Teachers by the Same Class

Class	No. of Stud.	Teacher-Student Relations			Grades and Testing			Materials			Personality			Methods		
		Mean Item Rating		t Value	Mean Item Rating		t Value	Mean Item Rating		t Value	Mean Item Rating		t Value	Mean Item Rating		t Value
		Teach. A	Teach. B		Teach. A	Teach. B		Teach. A	Teach. B		Teach. A	Teach. B		Teach. A	Teach. B	
1	13	4.14	3.96	0.76	4.13	4.39	-2.46*	4.23	4.18	0.49	4.26	3.72	3.41**	4.01	3.92	0.78
2	22	3.71	3.74	-0.19	4.27	3.83	2.87**	4.17	3.64	2.98**	3.42	4.26	-6.27**	3.86	3.12	4.98**
3	24	4.26	3.86	3.65**	4.17	4.03	0.94	4.49	4.00	3.61**	3.97	3.74	1.87	3.92	3.78	1.34
4	14	4.07	4.43	-1.84	4.22	4.48	-2.15	3.98	4.36	-2.58*	3.74	4.67	-5.64**	3.68	4.12	-4.89**
5	16	4.71	4.38	3.65**	4.50	3.61	6.63**	4.60	4.21	2.64*	4.50	4.50	0.00	4.34	3.92	2.49*
6	18	4.64	4.42	1.70	4.44	3.91	2.50*	4.46	4.41	0.41	4.54	4.57	-0.21	4.09	4.00	0.51
7	25	4.14	4.30	-1.28	3.99	4.23	-2.17*	4.00	4.29	-2.40*	4.20	4.36	-1.33	3.95	4.07	-1.42
8	17	3.74	4.00	-2.62*	3.57	4.10	-3.39**	3.67	4.04	-2.12*	3.78	4.04	-1.28	3.50	3.87	-3.24**
9	15	4.29	3.92	2.71*	4.50	3.72	3.69**	4.22	4.13	0.53	4.44	4.31	1.25	4.09	3.87	1.38

*Significant at beyond the .05 level

**Significant at beyond the .01 level

Table 6

Comparisons of Ratings Given the Same Teacher by Two Classes

Teacher	Number of Students		Teacher-Student Relations			Grades and Testing			Materials			Personality			Methods		
			Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating		
	Class 1	Class 2	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value
A	26	26	3.57	3.87	3.91	3.95	4.20	4.82*	3.38	4.04	12.89**	3.49	4.06	7.59**	3.43	3.76	6.00*
B	19	24	3.61	3.67	0.05	3.92	3.49	2.96	4.04	4.00	0.02	3.95	3.72	0.78	3.68	3.53	0.52
C	28	21	3.88	3.70	0.92	3.97	3.96	0.00	4.38	4.30	0.30	3.80	3.60	0.91	3.58	3.71	0.47
D	21	15	4.14	3.63	4.11	4.41	3.91	9.56**	4.35	4.18	0.80	4.22	3.76	6.23*	4.13	3.58	8.28**
E	23	14	3.48	3.31	0.72	4.39	4.44	0.10	4.48	4.48	0.00	3.48	3.45	0.01	3.90	3.67	1.32
F	9	20	4.02	3.90	0.36	4.41	4.19	1.24	4.22	3.92	2.14	4.00	4.03	0.01	3.60	3.81	0.64
G	26	27	4.20	4.66	13.15**	4.34	4.72	11.61**	4.06	4.38	3.73	4.58	4.89	12.56**	4.18	4.57	15.18**
H	19	15	4.49	4.60	0.44	4.63	4.75	1.20	4.32	4.38	0.12	4.49	4.62	0.50	4.04	4.08	0.09

*Significant at beyond the .05 level

**Significant at beyond the .01 level

Table 7

Correlations Between StOTT Ratings and Both Student and Teacher Characteristics (N = 334)[1]

Variable	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Student Grade Level (10,11,12)	.04	-.03	.03	.15**	.08
Student Sex (1 = Male, 2 = Female)	.18**	-.01	.11*	.19**	.03
Student Grade in Teacher's Course (A = 5, B = 4, etc.)	.27**	.05	.13**	.23**	.19**
Student Average Grade for Major Courses (A = 5, B = 4, etc.)	.19**	-.02	.13**	.16**	.03
Difference Between Student Grade in Course and Grade in Major Courses[2]	.16**	.08	.03	.14**	.20**
Student Reason for Taking Course (2 = Wanted To, 1 = Did Not Want To)	.09*	.11**	.07	.07	.17**
Teacher Sex (1 = Male, 2 = Female)	.23**	.07	-.02	.29**	.07

Table 8

Correlations Between Mean StOTT Ratings Given a Teacher and Both Student and Teacher Characteristics (N = 18)[1]

	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Student Grade Level	.20	-.06	-.06	.34	.21
Student Sex	.68**	-.07	.41*	.57*	.12
Student Grade in Teacher's Course	.63**	-.01	.31	.54**	.19
Student Average Grade for Major Courses	.60**	-.03	.21	.57**	.04
Difference Between Student Grade in Course and Grade in Major Courses[2]	.14	.02	.39	.04	.21
Student Reason for Taking Course	.03	.43*	.08	-.06	.35
Teacher Sex	.41*	.00	.09	.51*	.21

*Significant at beyond the .05 level
**Significant at beyond the .01 level

[1]Since hypotheses about the direction of differences were not stated for Student Grade Level, Student Sex and Teacher Sex, two-tailed tests were carried out for these. For all other variables one-tailed tests were performed.

[2]Positive values for this variable indicated that the teacher being rated gave students grades which were higher than those being received in their major subject courses.

Many of the correlations shown in Table 8 are significant and are of rather large magnitude. Thus, much of the reason for the finding of significant relationships for the total group appears to be that there was a tendency for classes which gave teachers high ratings to differ from classes which gave low ratings. Classes which gave high ratings tended to be made up of a higher proportion of female students and to include students who were earning higher grades both in the rated teacher's course and in their other courses. (The correlation between these two variables was .50.) Female teachers tended also to be rated higher than male teachers. Ratings which were significantly related to Teacher Sex and to student characteristics were limited almost exclusively to the Teacher-Student Relations subscale and the Personality subscale.

Table 9 contains correlations computed between student ratings and their scores on the Interest in School and Learning instrument. These two instruments were administered to eight classes of students.

Almost all of the correlations shown in Table 9 are significant at beyond the .05 level (one-tailed test). Students who held the most positive attitudes toward Learning, toward School and toward Teachers did tend to give teachers the highest StOTT ratings.

Table 9

Correlations Between StOTT Ratings and Interest in School and Learning Scores (N = 161)

Interest in School and Learning Subscale	Teach.-Stud. Relations	Grades/ Testing	Materials	Personality	Methods
Learning	.18**	.08	.30**	.04	.28**
School	.24**	.18**	.29**	.16*	.30**
Teachers	.35**	.33**	.34**	.23**	.40**
Total Instrument	.30**	.22**	.36**	.18**	.38**

*Significant at beyond the .05 level
**Significant at beyond the .01 level

Finally, Table 10 was constructed as a means of examining the effect of the subject area taught upon the ratings given a teacher. The number of teachers representing each subject area was felt to be too small for a statistical test to be performed. However, informal examination of Table 10 seems to indicate that the subject area taught did not have a major impact upon ratings. For subject areas in which a number of teachers were included, ratings varied rather widely as a function of individual teachers.

Table 10

Mean Item Ratings Obtained on Each Subscale by
All 36 Teachers in the Study

Subject Area	Teacher	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Mathematics	A	4.39	4.46	4.32	4.46	3.81
	B	4.55	4.70	4.35	4.57	4.07
	C	4.26	4.17	4.49	3.97	3.92
Social Science	A	4.36	3.98	4.44	4.06	4.05
	B	4.60	4.55	4.44	4.79	4.41
	C	3.72	4.08	3.71	3.78	3.60
	D	3.79	3.96	4.34	3.70	3.65
	E	4.43	4.53	4.22	4.74	4.38
	F	3.94	3.78	3.84	3.99	3.73
	G	4.24	3.74	4.25	4.46	3.93
	H	3.90	4.22	4.08	3.66	3.81
	I	3.86	3.88	4.59	3.75	3.61
Foreign Language	A	4.59	4.82	4.73	4.60	4.43
	B	4.37	4.62	4.31	4.51	4.18
English	A	4.22	4.33	4.37	4.56	4.08
	B	4.30	4.48	4.23	4.24	4.14
	C	4.26	4.22	4.05	4.40	4.03
	D	4.36	4.31	4.36	4.41	4.15
	E	4.55	4.48	4.43	4.49	4.18
	F	4.14	4.13	4.23	4.26	4.01
	G	4.43	4.48	4.36	4.67	4.12
Science	A	4.55	4.08	4.01	4.59	4.07
	B	3.64	3.71	4.02	3.84	3.61
	C	3.73	3.83	3.64	4.26	3.12
Business	A	4.69	4.78	4.68	4.37	3.84
	B	4.59	4.09	4.56	4.74	4.14
	C	3.96	4.30	4.07	4.02	3.71
Vocational Education	A	4.03	4.25	4.22	3.89	4.08
	B	4.01	4.43	4.41	4.04	3.95
	C	3.60	4.54	4.60	3.42	4.11
	D	3.58	4.13	4.33	3.22	3.73
	E	3.88	4.51	4.57	3.43	3.95
	F	3.21	4.25	3.94	3.72	3.89
	G	3.85	4.38	4.38	3.96	3.93
	H	3.88	4.16	4.27	3.99	3.86
	I	3.40	4.42	4.48	3.47	3.79

Factor Analysis of the StOTT

The present study provided empirical support for scoring the StOTT in terms of five subscales. The items included on a subscale also appeared to be highly similar in content.

However, it was believed possible that a factor analysis of the instrument might add measurably to the knowledge which had been gained about it. Therefore, item data from the total sample of 925 students were entered into a principal axis factor analysis. Squared multiple correlations were used as communality estimates and iterations were carried out until convergence had been achieved. Both varimax and oblique rotations were attempted. The results for these, in terms of factor structure, were almost identical.

Six factors with eigenvalues greater than 1.0 were extracted. Unfortunately the six-factor solution was not readily interpretable and included a single-item factor. This was also the case for the four- and five-factor solutions. The three-factor solution was interpretable. At least five items loaded highly on each factor. In addition, the scree test (Cattell, 1966) indicated that the three-factor solution was appropriate. As a means of determining whether or not this solution was a function of the particular factoring method used, the three-factor solution for a principal components analysis was also examined and found to be highly similar.

The three factors extracted accounted for a total of 40 per cent of the variance in the 29 items. The factors extracted were called:

I. Student Perception of the Teacher (8 items)--made up of six Teacher-Student Relations items and two Personality items.
II. Methods and Materials (16 items)--made up of all items included in Grades and Testing and in Materials and of most Methods items.
III. Teacher Ability to Interest Students (5 items)--made up of one Personality item (enthusiasm for the subject taught) and of four Methods items.

The first two factors extracted appeared to be more general in nature than are the StOTT subscales. In essence, they combined similar StOTT subscales. The third factor, however, seemed to provide a new dimension. The items loading highly on it all had to do with a teacher's ability to interest students. Items loading on this factor measured student opinions about the meaningfulness of a teacher's lectures and discussions, the teacher's enthusiasm for his or her subject, the teacher's ability to stimulate discussion and the extent to which the teacher varied his or her methods of presentation. Scores obtained on this factor should provide a valuable addition to those obtained on the five StOTT subscales.

Discussion and Conclusions

The Student Observation of Teachers and Teaching Techniques instrument appears to be a highly useful one for high school teachers. Student ratings of a teacher with the instrument were found to be stable over an interval of a month. Students seemed capable of making and repeating discriminations among the five scale points for each item. Classes also were able to employ the StOTT to pick out ways in which they felt their teachers differed. The subscales of the instrument make sense conceptually and, in general, were found to be internally consistent.

The major problem uncovered in the study of the instrument was that students of differing types tended to rate teachers differently. If the results of this part of the study are replicable (and, indeed, they are similar to those of Smith and Brown, 1976) it would be expected that higher StOTT ratings would be routinely received by teachers teaching students who like school and are doing well in it than by teachers teaching students who are "turned off" by school and are not doing well.

These findings are not surprising ones. Students who hold generally poor attitudes toward teachers, for example, would not be expected to give high ratings to individual ones.

What the results do seem to suggest is that student observation of teachers instruments should not be used to compare high school teachers or to evaluate them. The types of students being taught appear to influence ratings of teachers to too great an extent for this. The present study has demonstrated that it is even possible for two similar classes taking the same course from a teacher to disagree in their ratings.

These results should not be viewed as negative ones with respect to the use of the StOTT by teachers within their own classrooms. Teachers who are aware of the type of student they are teaching can take factors such as student attitudes into consideration when interpreting the ratings given them by a class.

The StOTT instrument, then, appears to be capable of fulfilling the purpose for which it was constructed. Its value lies in providing feedback to individual teachers about their students' opinions of their methods. The instrument was not intended to be a device for comparing the effectiveness of teachers. The results obtained in the present study indicate that it should not be used for this purpose.

References

Cattell, R. B. The scree test for the number of factors. Multivariate Behavioral Research, 1(2):245, 1966.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Davidoff, S. H. The development of an instrument designed to secure student assessment of teaching behaviors that correlate with objective measures of student achievement. Paper presented at the American Educational Research Association Annual Meeting, Minneapolis, 1970.

DeAmico, G. Second Year Report: A Student's Rating Scale of a Teacher. Sheboygan Public Schools, Wisconsin, 1973.

Hayes, R. B., Keim, F. N. and Neiman, A. M. The Effects of Student Reactions to Teaching Methods. USOE-Project No. 5-0096, 1966.

Hayes, R. B., Keim, F. N. and Neiman, A. M. The Effects of Student Reactions to Teaching Methods. USOE-Project No. 6-2056, 1967.

Hertzog, J. F. Getting Inside the EQA Inventory. Harrisburg: Pennsylvania Department of Education, 1976.

Jackson, M. L. and Fuller, F. F. Influence of social class on students' evaluations of their teachers. Proceedings of the 74th Annual Convention of the American Psychological Association, 1966.

Masters, J. R., Shannon, G. A. and Reardon, F. J. A Validation Study of the Educational Quality Assessment Interest in School and Learning Instruments. Harrisburg: Pennsylvania Department of Education, 1975.

Shaw, J. S. Students evaluate teachers and (better sit down) it works. Nation's Schools, Volume 91, No. 4, April 1973.

Smith, J. P. and Brown, T. J. Relationships among secondary students' evaluations of their courses and teachers and their general attitudes toward school. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, 1976.

Thompson, J. E. Student evaluation of teachers. NASSP Bulletin, October 1974.

Veldman, D. J. and Peck, R. F. The Pupil Observation Survey: Teacher Characteristics from the Students' Viewpoint. Austin, Texas: Texas University, Research and Development Center for Teacher Education, 1967.

Appendix

Student Name _____ Subject _____
School _____ Date _____
Grade _____ Period _____
Teacher Observed _____

STUDENT OBSERVATION OF TEACHERS
(Version 2)

This form asks some of the same questions you have just answered, but in a different way. Again, please answer them as honestly as possible.

DIRECTIONS: On the following page are pairs of statements. Use these pairs of statements to show what you have observed about your teacher. For each question, compare each of the two statements. Which of the two statements most describes your teacher? Choose the one which does and circle its letter in the Answer column.

Example	Answers
Which statement of each pair is most true of your teacher?	
1. A - The teacher gives meaningful homework. B - The teacher is humorous.	1. A or (B)
2. C - The teacher works hard. D - The teacher is helpful.	2. (C) or D
3. E - The teacher is neat. F - The teacher gives good lectures.	3. E or F

The student who answered question 1 above felt that statement B (The teacher is humorous.) described his or her teacher better than statement A (The teacher gives meaningful homework.).

For question 2, the student believed statement C (The teacher works hard.) was more true of his or her teacher than was statement D (The teacher is helpful.)

Which statement above, E or F, describes your teacher best? To answer question 3 you would circle either E or F in the column below the word Answers.

Continue in this way on the following page until you have answered all 16 questions.

Which statement of each pair is most true of your teacher?

		Answers
1.	F--The teacher varies manner of presentation. B--The teacher willingly accepts constructive criticisms from students.	1. F or B
2.	D--The teacher is fair and impartial in dealing with all students. E--The teacher stimulates discussions.	2. D or E
3.	A--The teacher is willing to help you when you need it. F--The teacher varies manner of presentation.	3. A or F
4.	D--The teacher is fair and impartial in dealing with all students. F--The teacher varies manner of presentation.	4. D or F
5.	H--The teacher comes to class well prepared. D--The teacher is fair and impartial in dealing with all students.	5. H or D
6.	E--The teacher stimulates discussions. C--The teacher values student opinions.	6. E or C
7.	C--The teacher values student opinions. H--The teacher comes to class well prepared.	7. C or H
8.	B--The teacher willingly accepts constructive criticisms from students. G--The teacher remains on one subject area until most students understand it.	8. B or G
9.	E--The teacher stimulates discussions. B--The teacher willingly accepts constructive criticisms from students.	9. E or B
10.	E--The teacher stimulates discussions. A--The teacher is willing to help you when you need it.	10. E or A
11.	F--The teacher varies manner of presentation. C--The teacher values student opinions.	11. F or C
12.	C--The teacher values student opinions. G--The teacher remains on one subject area until most students understand it.	12. C or G
13.	G--The teacher remains on one subject area until most students understand it. A--The teacher is willing to help you when you need it.	13. G or A
14.	H--The teacher comes to class well prepared. B--The teacher willingly accepts constructive criticisms from students.	14. H or B
15.	A--The teacher is willing to help you when you need it. H--The teacher comes to class well prepared.	15. A or H
16.	D--The teacher is fair and impartial in dealing with all students. G--The teacher remains on one subject area until most students understand it.	16. D or G